Nonparametric Estimation of the Fractional Derivative of a Distribution Function

Andreea Borla, Costin Protopopescu

GREQAM
Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes études en Sciences Sociales
Universités d'Aix-Marseille II et III

Document de Travail
n°2010-45

DT-GREQAM

Nonparametric Estimation of the Fractional Derivative of a Distribution Function

Andreea Borla
Costin Protopopescu

November 2010

Nonparametric Estimation of the Fractional Derivative of a Distribution Function

by Andreea BORLA and Costin PROTOPOPESCU

Université de la Méditerranée, and GREQAM
Centre de la Vieille Charité
2, rue de la Charité
13002 Marseille FRANCE
borla@univmed.fr
protopopescu@univmed.fr

This version: May 2010

Abstract

We propose an estimator for the α fractional derivative of a distribution function. Our estimator, based on finite differences of the empirical df generalizes the estimator proposed by Maltz (1974) for the nonnegative real case. The asymptotic bias, variance and the consistency of the estimator are studied. Finally, the optimal choice for the "smoothing parameter" proves that even in the fractional case, the Stone's rate of convergence is achieved.

Keywords: fractional derivative, nonparametric estimation, distribution function, generalized differences

AMS Subject classification: Primary 62G05, Secondary: 26A33

1 Introduction

From the conceptual point of view, historically, it appears that a number of theoretical results in many different disciplines were obtained by considering only the integer order derivatives. In consequence, the statistical analysis of the correspondent models was limited to estimate (by direct or indirect inference) only "classical" derivatives as parameters of interest.

On the other hand, the evolution of knowledge in all these domains, pushed the theoretical studies to emphasize models dealing with the concept of fractional derivative. In fact, note that this concept is not a recent one at all, it dates back to 1695 when, in a letter to L'Hôpital, raised the question if the meaning of derivatives with integer order can be generalized to derivatives with non-integer orders. This question was an ongoing topic for more than 300 years and mathematicians such as Lagrange, Laplace, Fourier, Liouville, Riemann, Weyl, Abel, Lacroix, Leibnitz brought their co ntribution to this field. For a historical survey, see [12]. For a comprehensive introduction to fractional derivatives see [18].

In the last twenty years, fractional calculus operators were used in different statistical contexts: computation of the fractional moments of distributions ([6]), finding the exact distribution of the Wald statistic, or the SUR estimator ([14] and [15]), deriving the moments of OLS and 2SLS estimators ([5]); continuous time random walks and its applications in finance ([19]), the fractional brownian motion and its applications in finance ([1] and [7]). We can also mention numerous developments in mechanics and engineering based on the fractional differential equations ([8], [16]), in physics ([4] and [10]), nonlinear dynamics ([3] and [11]).

In this paper, we will use a generalization of the numerical derivative operator in order to give an estimator for the fractional derivative of a distribution function. This allows a broad range of applications where the parameters of interest are in particular, densities or distribution functions. Despite the fact that the mathematical tools needed for this development were available and in the literature enough applications of the concept appeared, the estimator of the fractional derivative was never considered from a statistical point of view.

2 Nonparametric estimators of the derivatives

Let X_1, X_2 ... X_n be a random sample distributed according to a distribution function $\mathbf{F}$. [9] gave an estimator for the k^{th} derivative $\mathbf{F}^{(k)}$ at a point x, where k is a natural number. This estimator is based on a symmetrized finite difference approximation of the derivative of the function:

$$\widetilde{\mathbf{F}}_n^{(k)}(x,h) = (2h)^{-k} \sum_{j=0}^{k} (-1)^j \widehat{\mathbf{F}}_n(\widetilde{x}_j) \binom{k}{j} \tag{1}$$

where the knots are $\widetilde{x}_j = x + (k-2j)h$, $\widehat{\mathbf{F}}_n$ is the empirical distribution function based on the random sample, and $h = h_n$ is a sequence of positive numbers converging to zero.

In his paper, Maltz proved the following results:

Theorem 1. *Assume that $\mathbf{F}^{(k)}$ (or $\mathbf{F}^{(k+2)}$) exists at x. Then:*

$$\begin{aligned}\mathbf{E}\left[\widetilde{\mathbf{F}}_n^{(k)}(x,h)\right] &= \mathbf{F}^{(k)}(x)+o(1)\\ &= \mathbf{F}^{(k)}(x)+\frac{k}{6}h^2\mathbf{F}^{(k+2)}(x)+o(h^2)\end{aligned}$$

Proof. See [9]. □

Theorem 2. *Assume that $\mathbf{F}^{(1)}$ exists at x. Then:*

$$\mathbf{Var}\left[\widetilde{\mathbf{F}}_n^{(k)}(x,h)\right]=n^{-1}(2h)^{1-2k}\mathbf{F}^{(1)}(x)\binom{2k-2}{k-1}+o(n^{-1}h^{1-2k})$$

Proof. See [9]. □

Their immediate consequences are the following corollaries:

Corrolary 1. *If $\mathbf{F}^{(k)}$ exists at x and the sequence $\{h_n\}$ satisfies the conditions $h_n\to 0$ and $nh_n^{2k-1}\to\infty$ as $n\to\infty$ then $\widetilde{\mathbf{F}}_n^{(k)}(x,h)\xrightarrow[n\to\infty]{\mathbf{P}}\mathbf{F}^{(k)}(x)$.*

Corrolary 2. *If $\mathbf{F}^{(k+2)}$ exists at x, then the mean square error (MSE) of $\widetilde{\mathbf{F}}_n^{(k)}(x,h)$ is given by:*

$$\begin{aligned}\mathbf{MSE}\left[\widetilde{\mathbf{F}}_n^{(k)}(x,h)\right] &= \frac{k^2h^2}{36}\left[\mathbf{F}^{(k+2)}\right]^2+n^{-1}(2h)^{1-2k}\mathbf{F}^{(1)}(x)\binom{2k-2}{k-1}\\ &\quad +o(h^4+n^{-1}(2h)^{1-2k})\end{aligned}$$

This estimator has the same rate of convergence as the "smoothed" version based on the Parzen-Rosenblatt kernel, that is of order $\sqrt{nh^{2k-1}}$. Moreover, the "optimal" order (in the sense of minimizing the MSE) of the "smoothing parameter" is proportional to $n^{-\frac{1}{2k+3}}$.

Before the development of the definition of our estimator, let us make some remarks about Maltz's estimator.

First, note that (1) uses only differences between the values of the empirical distribution (which is not a differentiable function) in order to compute an approximation for the k derivative of the true distribution function $\mathbf{F}$. The estimations based on simulated data proved us that it is a powerful nonparametric estimator which can reveal very well the structure of a derivative. Just like the classical smoothed version of Parzen-Rosenblatt based on a kernel, it has a smoothing parameter h which may be assimilated to a classical "bandwidth" making the trade-off between the bias and the variance.

Moreover, the first order derivative computed with Maltz's formula coincides with the Parzen-Rosenblatt density estimator with an uniform kernel. So the smoothing parameter has the same interpretation in both estimators.

All these facts convinced us to generalize this estimator to the fractional case, given that in the literature of numerical analysis there exists an equivalent for the numerical approximation of the fractional derivative based on generalized differences.

We start by taking a look at the estimator as functional operator acting on $\mathbf{F}$. The key role of the paper will be played by the operator $\overline{\mathbf{\Delta}}_\delta^k$, defined as $\sum_{i=0}^k (-1)^i \binom{k}{i} \mathbf{F}(x + (k/2 - i)\delta)$, which performs the transformation used in the definition (1) of the estimator studied by Maltz. In the literature of numerical analysis it is called the *symmetric δ-shifted difference operator of order k*, because of the "shifted" knots $x + (k/2 - i)\delta$ used as arguments for the function $\mathbf{F}$.

From the numerical perspective, this class of operators is intimately related to the approximation of the classical k^{th} derivative operator $\mathbf{F}^{(k)}$, because we can find a functional norm that verifies $\left\| \left[\overline{\mathbf{\Delta}}_\delta^k \mathbf{F}\right](x) - \delta^k \mathbf{F}^{(k)} \right\| < \eta(\delta)$.

More exactly, for every $x \in supp(F)$ (such that. $\mathbf{F}^{(1)}(x) \neq 0$), there is an order r such that :

$$\mathbf{F}^{(k)}(x) = \delta^{-k} \left[\overline{\mathbf{\Delta}}_\delta^k \mathbf{F}\right](x) + O(\delta^r)$$

Using the translation operator τ_δ defined by $[\tau_\delta \mathbf{F}](x) = \mathbf{F}(x + \delta)$, if we look at the estimator (1), we can write:

$$\widetilde{\mathbf{F}}_n^{(k)}(x,h) \stackrel{def}{=} (2h)^{-k} (\overline{\mathbf{\Delta}}_{2h}^k \widehat{\mathbf{F}}_n)(x) = \delta^{-k} (\tau_\delta - 1)^k \tau_{-\delta/2}^k \widehat{\mathbf{F}}_n(x)$$

for $\delta = 2h$, here $\widehat{\mathbf{F}}_n$ being the empirical distribution function. This is not only a numerical approximation for the k^{th} derivative for $\mathbf{F}$. Given the random character of $\widehat{\mathbf{F}}_n$, Maltz proved that it is a very appealing nonparametric estimator for $\mathbf{F}^{(k)}$ too, which has all the classical properties desired by statisticians. The strong convergence of $\widetilde{\mathbf{F}}_n^{(k)}$ to $\mathbf{F}^{(k)}$ comes from translation of the convergence of $\widehat{\mathbf{F}}_n$ (a $\sqrt{n}$-convergent estimator for the true cumulative distribution function $\mathbf{F}$) made by the continuous linear operator $\overline{\mathbf{\Delta}}_\delta^k$.

The next step is to take Maltz's definition that works only for the integer values of the order k, and to generalize it to a non-integer order.

Definition 1. *For a positive scalar α, and a real function $\mathbf{F}$, let us define $\overline{\mathbf{\Delta}}_\delta^\alpha$ as the class of* generalized symmetric δ-shifted difference operator *indexed by some positive parameter δ:*

$$\left[\overline{\mathbf{\Delta}}_\delta^\alpha \mathbf{F}\right](x) \stackrel{def}{=} (\tau_\delta - 1)^\alpha \tau_{-\delta/2}^\alpha \mathbf{F}(x) \tag{2}$$

where $(\tau_{-\delta} - 1)^\alpha$ is computed using the generalized binomial formula.

In a more explicit way, we can write:

$$\left[\overline{\mathbf{\Delta}}_\delta^\alpha \mathbf{F}\right](x) = \sum_{l \geq 0} (-1)^l \binom{\alpha}{l} \mathbf{F}(x + (\delta/2 - l)\delta)$$

The binomial coefficients $\binom{\alpha}{l}$ are seen in the generalized sense $\dfrac{\Gamma(\alpha+1)}{l!\Gamma(\alpha - l + 1)}$ and involve the Gamma function (which extends the classical factorial function). It is easy to prove that for any bounded function $\mathbf{F}$, this infinite alternate series is convergent.

Remark: If α is a positive integer then all the binomial coefficients for $l > \alpha$ become zero and we have Maltz's estimator as a special case.

This formula defines a natural approximation for the derivative of order α as a generalization of the integer case, and its limit when δ goes to 0 is called *the Grünwald-Letnikov derivative,* denoted here by $\mathbf{F}^{(\alpha)}(x)$.

In the same spirit as below,given x, for each α it exists some positive integer m for which locally:

$$\delta^{-\alpha}\left[\overline{\boldsymbol{\Delta}}_{\delta}^{\alpha}\mathbf{F}\right](x)=\mathbf{F}^{(\alpha)}(x)+O\left(\delta^{m}\right)$$

This is a natural generalization of a classical result involving numerical approximation of the true derivative of integer order.

Using the plug-in technique, in order to obtain a nonparametric estimator of the fractional derivative of a distribution function $\mathbf{F}$, we propose to replace in (2) the unknown $\mathbf{F}$ by its nonparametric estimator $\widehat{\mathbf{F}}_n$.

Definition 2. *Given a positive scalar α, the estimator of the $\alpha-$fractional derivative of a distribution function $\mathbf{F}$ is given by:*

$$\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\stackrel{def}{=}\delta^{-\alpha}\left(\tau_{\delta}-1\right)^{\alpha}\tau_{-\delta/2}^{\alpha}\widehat{\mathbf{F}}_n(x) \tag{3}$$

where the parameter $\delta=\delta_n$ is a sequence of positive numbers converging to zero as $n\rightarrow\infty$.

Other conditions will be imposed on δ in the next sections.

This estimator will be studied in the remaining of the paper. The next section deals with the asymptotic bias, section 3 presents the asymptotic variance. The convergence is proven in section 4. Finally, the last section outlines the optimal choice of the smoothing parameter.

3 Asymptotic Bias

The derivation of the asymptotic bias cannot be carried out by the usual technique applied in the case of nonparametric estimators, in particular the one Maltz used in his paper. In order to compute the bias, he performs a Taylor development of order k around x, obtains a polynomial formula with respect to δ and uses some identities involving binomial coefficients to obtain the result (see his paper for details). We cannot use this simple technique directly, because there is no equivalent of the Taylor development of a fractional order.

Instead, there exist some results used in the context of the numerical analysis for computing the extrapolation to the limit for the numerical fractional differentiation (see [20]), which provide a Taylor-like development for the differential operators, and which follow the same lines as our proof. Roughly speaking, using the Fourier transform applied to some functionals, we can deduce an equivalent Taylor expansion for the numerical differential operators. However, these results are written for general functions lying in $\mathbb{R}$ and which don't have the behaviour of a distribution function. Consequently, necessary conditions must be imposed in order to prove a Taylor-like development for the class of shifted generalised difference operators Δ_{δ}^{α}. More exactly, $\mathbf{F}$ must verify some very restrictive regularity conditions (the classical derivatives of order up to $[\alpha]+6$ must be in $\mathbf{L}^1(\mathbb{R})$) which is a drawback in our context, unless the probability distribution function verifies $\mathbf{F}\in C^{\infty}(\mathbb{R})$ (like the Gaussian, Student or Gamma distribution etc).

For any $\mathbf{g}$ in $\mathbf{L}^1(R)$, the space of absolute integrable functions, we consider the following definition for the Fourier transform of g :

$$\mathcal{F}(\mathbf{g})(x) \stackrel{def}{=} \int_{\mathbb{R}} \mathbf{g}(t) \exp(itx) dt$$

and can be extended to the generalised shifted central difference operator.

We will use it in order to prove that the bias of the fractional estimator has the same order of magnitude with respect to δ as the one for integer case, i.e. is of order $O(\delta^2)$ under some regularity conditions on F.

Theorem 3. *If the true cumulative distribution function* $\mathbf{F} \in \mathcal{C}^{[\alpha]+5}(\mathbf{R})$ *and* $\mathbf{F}^{(s)} \in \mathbf{L}^1(\mathbb{R})$ *for all integer* s *up to* $[\alpha]+6$*, then the asymptotic bias of the estimator (3) is given by:*

$$\mathbf{B}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right] = \frac{1}{4!}\alpha\delta^2\mathbf{F}^{(\alpha+2)}(x) + o(\delta^2) \tag{4}$$

Proof. Let's start by writing the mean of the estimator. For any x in the support of $\mathbf{F}$ we have:

$$\mathbf{E}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right] = \mathbf{E}\left[\delta^{-\alpha}\left(\tau_\delta - 1\right)^\alpha \tau_{-\delta/2}^\alpha \widehat{\mathbf{F}}_n(x)\right]$$

The linearity of the expectation operator $\mathbf{E}$ and respectively of the translation operator τ_δ gives:

$$\begin{aligned}\mathbf{E}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right] &= \delta^{-\alpha}\left(\tau_\delta - 1\right)^\alpha \tau_{-\delta/2}^\alpha \mathbf{E}\left[\widehat{\mathbf{F}}_n(x)\right] \\ &= \delta^{-\alpha}\left(\tau_\delta - 1\right)^\alpha \tau_{-\delta/2}^\alpha \mathbf{F}(x)\end{aligned}$$

In an equivalent manner, using the generalised symmetric δ-shifted difference operator defined in (2) we can write:

$$\mathbf{E}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta_n)\right] = \delta^{-\alpha}(\overline{\mathbf{\Delta}}_\delta^\alpha \mathbf{F})(x)$$

We can easily show that [1]:

$$\mathcal{F}(\overline{\mathbf{\Delta}}_\delta^\alpha \mathbf{F})(t) = (-it\delta)^\alpha \left[\frac{\sin(\frac{t\delta}{2})}{\frac{t\delta}{2}}\right]^\alpha \mathcal{F}(\mathbf{F})(t)$$

Now for all $u \neq 0$, plugging the Taylor series representation for $\dfrac{\sin(u)}{u}$ in the previous equation we find:

$$\delta^{-\alpha}\mathcal{F}(\overline{\mathbf{\Delta}}_\delta^\alpha \mathbf{F})(t) = (-it)^\alpha \left[\sum_{k=0}^{\infty} \frac{(-1)^k (\frac{t\delta}{2})^{2k}}{(2k+1)!}\right]^\alpha \mathcal{F}(\mathbf{F})(t).$$

Formally, given that the series is convergent, by the multinomial formula that generalises Newton's binomial theorem, it can be written as :

$$\delta^{-\alpha}\mathcal{F}(\overline{\mathbf{\Delta}}_\delta^\alpha \mathbf{F})(t) = (-it)^\alpha \left[b_0 + b_2(t\delta)^2 + O(t^4\delta^4)\right] \mathcal{F}(\mathbf{F})(t) \tag{5a}$$

[1] Remember that here the cumulative distribution $\mathbf{F} \notin \mathbf{L}^1(\mathbb{R})$ but can be seen as a tempered distribution, so it has a Fourier transform.

We are only interested in the coefficients of order up to 2. By identification of the coefficients we have $b_0 = 1$ and $b_2 = \frac{1}{4!}\alpha$, the other computable coefficients b_{2k} are not relevant for our proposal, being stored into the $O(\cdot)$ term.

Let us define a reminder $\mathbf{r}_2(t)$ ensuring that we can rewrite the previous result as:

$$\begin{aligned}\delta^{-\alpha}\mathcal{F}(\overline{\mathbf{\Delta}}^{\alpha}_{\delta}\mathbf{F})(t) &= (-it)^{\alpha}\left[\sum_{k=0}^{1} b_{2k}(t\delta)^{2k} + \sum_{k=2}^{\infty} b_{2k}(t\delta)^{2k}\right]\mathcal{F}(\mathbf{F})(t) \\ &= (-it)^{\alpha}\left[b_0 + b_2(t\delta)^2\right]\mathcal{F}(\mathbf{F})(t) + (-it)^{\alpha}\mathbf{r}_2(t)\mathcal{F}(\mathbf{F})(t)\end{aligned}$$

and let ρ be the inverse Fourier Transform of $(-it)^{\alpha}\mathbf{r}_2(x)\mathcal{F}(\mathbf{F})(t)$, i.e.:

$$\mathcal{F}(\rho)(t,\delta) = (-it)^{\alpha}\left(\sum_{k=2}^{\infty} b_{2k}(t\delta)^{2k}\right)\mathcal{F}(\mathbf{F})(t)$$

Using the classical property of the Fourier transform:

$$(-it)^{\alpha+k}\ \mathcal{F}(\mathbf{F})(t) = \mathcal{F}\left(\mathbf{F}^{(\alpha+k)}\right)(t) \qquad k = 0,1,2$$

this allows us to write that:

$$\delta^{-\alpha}\mathcal{F}\left(\overline{\mathbf{\Delta}}^{\alpha}_{\delta}\mathbf{F}\right)(t) = \mathcal{F}\left(\mathbf{F}^{(\alpha)}\right)(t) + b_2\mathcal{F}\left(\mathbf{F}^{(\alpha+2)}\right)(t)\delta^2 + \mathcal{F}(\rho)(t,\delta) \tag{6}$$

After applying the inverse Fourier transform to the above identity, we have:

$$\delta^{-\alpha}\left(\overline{\mathbf{\Delta}}^{\alpha}_{\delta}\mathbf{F}\right)(x) = \mathbf{F}^{(\alpha)}(x) + \frac{1}{4!}\alpha\delta^2\mathbf{F}^{(\alpha+2)}(x) + \rho(x,\delta)$$

provided that $\mathcal{F}(\rho)(t,\delta)$ is in $\mathbf{L}^1(\mathbb{R})$. We will see in the following lines that this condition is verified. Moreover, in order to prove the result (4), we need to show that $\rho(x,\delta) = o\left(\delta^2\right)$.

First, observe that:

$$\mathcal{F}(\rho)(t,\delta) = (-it)^{\alpha}O(t^4\delta^4)\mathcal{F}(\mathbf{F})(t) = O(\delta^4)(-i)^{\alpha}t^{\alpha+4}\mathcal{F}(\mathbf{F})(t)$$

Under the assumptions on the derivatives of $\mathbf{F}$, one can prove that $(1+|t|)^{[\alpha]+6}\,|\mathcal{F}(\mathbf{F})(t)|$ is bounded by some constant C_1. Implicitly, after some simple algebraic calculus we deduce that $[\mathcal{F}(\rho)]\,(t,\delta)$ is in $\mathbf{L}^1(\mathbb{R})$. Applying the inverse Fourier transform, we obtain that $|\rho(x,\delta)| \le \widetilde{C}\delta^4$ i.e. $\rho(x,\delta) = o\left(\delta^2\right)$.

Thus, if we retain the terms up to $O(\delta^2)$ we have shown that:

$$\mathbf{E}\left[\widetilde{\mathbf{F}}^{(\alpha)}_n(x,\delta)\right] = \mathbf{F}^{(\alpha)}(x) + \frac{1}{4!}\alpha\delta^2\mathbf{F}^{(\alpha+2)}(x) + o(\delta^2)$$

and the theorem follows. □

4 Asymptotic variance

In our framework, the parameter δ can be seen as a smoothing parameter, just like the bandwidth h for the Parzen-Rosenblatt kernel estimator. Moreover, the computation of the variance will emphasise its dependence on the order of magnitude of δ and his exponent. This will affect the rate of convergence of our estimator, but given that the fractional derivative is a generalization of the standard concept of derivative, we will find a theoretical result which will act like a generalization of Maltz's and respectively Parzen-Rosenblatt's results.

Indeed, for the first order derivative of $\mathbf{F}$, [13] and [17] proved that the asymptotic variance of the estimator of $\mathbf{f}(\mathbf{x}) = \mathbf{F}^{(1)}(x)$, under the appropriate conditions on h_n is:

$$\lim_{n\to\infty}(nh)\mathbf{Var}(\widehat{\mathbf{f}}_n(x)) = \mathbf{f}(\mathbf{x})\int \mathbf{k}^2(u)du = \lambda_1\mathbf{F}^{(1)}(x)$$

On the other hand, Maltz pointed out that the estimator he proposed, under the appropriate conditions on h_n, verifies:

$$\lim_{n\to\infty}(nh^{2k-1})\mathbf{Var}(\widetilde{\mathbf{F}}_n^{(k)}(x,h)) = \mathbf{F}^{(1)}(x)\binom{2k-2}{k-1}$$

In the following we will generalise these two results to the fractional case. We begin by stating the main result and then we will prove it.

Theorem 4. *Assume that $\mathbf{F}^{(1)}(x)$ exists and $\delta = \delta_n$ is a sequence converging to zero which verifies $\lim\limits_{n\to\infty} n\delta^{2\alpha-1} = \infty$. Then for $\alpha > 1/2$:*

$$\mathbf{Var}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right] = n^{-1}\delta^{1-2\alpha}\mathbf{F}^{(1)}(x)\frac{\mathbf{\Gamma}(2\alpha-1)}{\left[\mathbf{\Gamma}(\alpha)\right]^2} + o(n^{-1}\delta^{1-2\alpha})$$

Proof. We have successively

$$\begin{aligned}
&\delta^{2\alpha}\mathbf{Var}\left(\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right) = \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{i+j}\binom{\alpha}{i}\binom{\alpha}{j}\mathbf{Cov}\left(\widehat{\mathbf{F}}_n(\widetilde{x}_i),\widehat{\mathbf{F}}_n(\widetilde{x}_j)\right)\\
&= n^{-1}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{i+j}\binom{\alpha}{i}\binom{\alpha}{j}\left[\mathbf{F}(\widetilde{x}_i\wedge\widetilde{x}_j) - \mathbf{F}(\widetilde{x}_i)\mathbf{F}(\widetilde{x}_j)\right]
\end{aligned}$$

where $\widetilde{x}_i \wedge \widetilde{x}_j$ denotes the minimum of $\widetilde{x}_i$ and $\widetilde{x}_j$. Now, assuming that $\mathbf{F}'(x)$ exists, by applying Taylor's rule we get:

$$\mathbf{F}(\widetilde{x}_i) = \mathbf{F}(x + (\alpha/2 - i)\delta) = \mathbf{F}(x) + (\alpha/2 - i)\delta\mathbf{F}^{(1)}(x) + o(\delta)$$

as $\delta \to 0$. Consequently:

$$\begin{aligned}
&\delta^{2\alpha}\mathbf{Var}\left(\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right) = \\
&= n^{-1}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{i+j}\binom{\alpha}{i}\binom{\alpha}{j}\Big[\mathbf{F}(x) - \mathbf{F}^2(x) + \\
&\{(\alpha/2 - i)\wedge(\alpha/2 - j)\}\,\delta\mathbf{F}'(x) - (\alpha - i - j)\delta\mathbf{F}(x)\mathbf{F}^{(1)}(x) + o(\delta)\Big]
\end{aligned} \tag{7}$$

The identity $\sum_{i=0}^{\infty}(-1)^i\binom{\alpha}{i} = 0$ will reduce the right part to : $\delta^{2\alpha}\mathbf{Var}\left(\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right) = -n^{-1}\mathbf{F}'(x)\delta\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{i+j}\binom{\alpha}{i}\binom{\alpha}{j}(i\vee j) + o(n^{-1}\delta)$, where $i \vee j$ denotes the maximum of i and j. Let's compute the double sum from (7), which will be a function depending on the value of α :

$$\mathbf{S}(\alpha) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{i+j}\binom{\alpha}{i}\binom{\alpha}{j}(i \vee j)$$

After some algebraic calculus, the same previous identity allows us to obtain:

$$\mathbf{S}(\alpha) = \sum_{i=0}^{\infty}(-1)^{i}\binom{\alpha}{i}[i\sum_{j=0}^{i}(-1)^{j}\binom{\alpha}{j} - \sum_{j=0}^{i}(-1)^{j}\binom{\alpha}{j}j]$$

On the other hand, the classical identities relative to the generalised binomial coefficients simplify even more the double series as being the difference between two simple series:

$$\begin{aligned}\mathbf{S}(\alpha) &= \sum_{i=0}^{\infty}\left[i(1+i)\binom{\alpha}{i}\binom{\alpha}{1+i}\left\{\frac{1}{\alpha} - \frac{1}{\alpha-1}\right\}\right] \\ &= -\frac{1}{\alpha(\alpha-1)}[\alpha\sum_{i=0}^{\infty}i\binom{\alpha}{i}^{2} - \sum_{i=0}^{\infty}i^{2}\binom{\alpha}{i}^{2}]\end{aligned}$$

We can decompose the previous sum in two parts: $\mathbf{S}_1(\alpha) = \sum_{i=0}^{\infty} i\binom{\alpha}{i}^2$ and respectively $\mathbf{S}_2(\alpha) = \sum_{i=0}^{\infty} i^2\binom{\alpha}{i}^2$. Both of them are functional series and the parameter α determines wether each of the two series is convergent or divergent. In order to study the convergence of the two previous series we use Kummer's test (see [2]). It is easy to see that:

$$\lim_{n\to\infty}(n+1)^{-2}[n^{3}\binom{\alpha}{n}^{2}\binom{\alpha}{n+1}^{-2} - (n+1)^{3}] = 2\alpha - 1$$

so we can conclude that $\mathbf{S}_2(\alpha)$ will be divergent or convergent as $2\alpha-1$ will be negative or positive. So for $0 < \alpha \leq 1/2$, $\mathbf{S}_2(\alpha)$ is divergent, and for $\alpha > 1/2$, $\mathbf{S}_2(\alpha)$ is convergent. Using the same criterion, we can verify that $\mathbf{S}_1(\alpha)$ is convergent for $\alpha > 0$ (the correspondent limit is infinite). On the other hand, for any $\alpha > 0$, we have the identity:

$$\sum_{i=0}^{\infty} i\binom{\alpha}{i}^{2} = \frac{\alpha^{2}\mathbf{\Gamma}(2\alpha)}{\mathbf{\Gamma}^{2}(\alpha+1)}$$

and

$$\sum_{i=0}^{\infty} i^{2}\binom{\alpha}{i}^{2} = \frac{\alpha^{2}\mathbf{\Gamma}(2\alpha-1)}{\mathbf{\Gamma}^{2}(\alpha)} \text{ for any } \alpha > 1/2$$

these identities being generalizations of the equivalent identities for the classical binomial coefficients extended to the Gamma function. In conclusion, if $\alpha > 1/2$, both series are convergent and using the previous results we can write:

$$\mathbf{S}(\alpha) = -\frac{1}{\alpha(\alpha-1)}\left[\frac{\alpha\mathbf{\Gamma}(2\alpha)}{\mathbf{\Gamma}^{2}(\alpha)} - \frac{\alpha^{2}\mathbf{\Gamma}(2\alpha-1)}{\mathbf{\Gamma}^{2}(\alpha)}\right] = -\frac{\mathbf{\Gamma}(2\alpha-1)}{\mathbf{\Gamma}^{2}(\alpha)}$$

so the variance of the estimator is:

$$\delta^{2\alpha}\mathbf{Var}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right] = n^{-1}\delta\mathbf{F}'(x)\frac{\mathbf{\Gamma}(2\alpha-1)}{[\mathbf{\Gamma}(\alpha)]^2} + o(n^{-1}\delta)$$

and the theorem follows. □

This result points out that the rate of convergence of our estimator will be of order $\sqrt{n\delta^{2\alpha-1}}$. For $\alpha = 1$ we find the classical result related to the kernel density estimator with uniform kernel. For an integer α we find Maltz's result.

The next section will briefly discuss the convergence in probability of the estimator.

5 Convergence

Theorem 5. *If $\mathbf{F}^{(\alpha)}$ exists at x and the sequence δ_n satisfies the conditions $\delta_n \to 0$ and $n\delta_n^{2\alpha-1} \to \infty$ as $n \to \infty$ then for all x in the support of the distribution $\mathbf{F}$ we have:*

$$\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta) \xrightarrow[n\to\infty]{\mathbf{P}} \mathbf{F}^{(\alpha)}(x)$$

Proof. Using Chebyshev's Inequality, we can write:

$$\Pr(\left|\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta_n) - \mathbf{E}\left(\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta_n)\right)\right| < \frac{\varepsilon}{2}) > 1 - \frac{4\mathbf{Var}\left(\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right)}{\varepsilon^2}$$

With the mean and the variance we computed before, we have :

$$\Pr(\left|\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta) - \mathbf{F}^{(\alpha)}(x) - \tfrac{1}{4!}\alpha\delta^2\mathbf{F}^{(\alpha+2)}(x) - o(\delta^2)\right| < \tfrac{\varepsilon}{2}) >$$
$$> 1 - \frac{4F'(x)\frac{\mathbf{\Gamma}(2\alpha-1)}{[\mathbf{\Gamma}(\alpha)]^2} + o(n^{-1}\delta^{1-2\alpha})}{n\delta^{2\alpha-1}\varepsilon^2}$$

Let us re-write the inequality between the brackets:

$$-\frac{\varepsilon}{2} - \frac{1}{4!}\alpha\delta^2\mathbf{F}^{(\alpha+2)}(x) + o(\delta^2) < \widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta) - \mathbf{F}^{(\alpha)}(x)$$
$$< \frac{\varepsilon}{2} - \frac{1}{4!}\alpha\delta^2\mathbf{F}^{(\alpha+2)}(x) + o(\delta^2)$$

Given that δ_n is a sequence converging to zero, one can find a rank $n_\varepsilon \in \mathbf{N}$ such that for all $n \geq n_\varepsilon$ we have:

$$\left|\frac{1}{4!}\alpha\delta^2\mathbf{F}^{(\alpha+2)}(x) - o(\delta^2)\right| < \frac{\varepsilon}{2}$$

So we can write respectively :

$$\Pr(\left|\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta_n) - \mathbf{F}^{(\alpha)}(x)\right| < \varepsilon) > 1 - \frac{4F'(x)\frac{\mathbf{\Gamma}(2\alpha-1)}{[\mathbf{\Gamma}(\alpha)]^2} + o(n^{-1}\delta^{1-2\alpha})}{n\delta^{2\alpha-1}\varepsilon^2}$$

or equivalently,

$$\Pr(\left|\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta) - \mathbf{F}^{(\alpha)}(x)\right| < \varepsilon) > 1 - c_n(x,\alpha)$$

where $\underset{n\to\infty}{lim}\, c_n(x,\alpha) = 0$, and the convergence in probability follows. □

6 Mean Square Error and the optimal choice of δ_n

Let's study the asymptotic behaviour of the estimator in terms of its mean square error. Using the precedent theorems, we can give the following result:

Theorem 6. *If $\mathbf{F}^{(1)}(x)$ and $\mathbf{F}^{(\alpha+2)}$ exists, for $\alpha > 1/2$ the asymptotic mean square error is given by:*

$$\mathbf{MSE}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right] = n^{-1}\delta^{1-2\alpha}\mathbf{F}^{(1)}(x)\frac{\mathbf{\Gamma}(2\alpha-1)}{[\mathbf{\Gamma}(\alpha)]^2} + \frac{1}{24^2}\alpha^2\delta^4\left[\mathbf{F}^{(\alpha+2)}(x)\right]^2 + o(\delta^4+n^{-1}\delta^{1-2\alpha})$$

Proof. The result is obtained by direct computation using the previous results from Theorem 3 and Theorem 4. □

Under the assumptions (i) $\delta \to 0$ as $n \to \infty$ (ii) $n\delta^{2\alpha-1} \to \infty$ as $n \to \infty$ we have that $\mathbf{MSE}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right] \to 0$ as $n \to \infty$.

Note that we have the same pattern as for the kernel estimation, our parameter acts in the same way as the bandwidth. If we choose a very small δ_n then the bias will be small but the variance will be large. A larger δ_n will give a smaller variance but a larger bias. Thus δ must be chosen in order to achieve the best trade-off between the two. We can choose an "optimal" smoothing parameter δ_n in order to minimise the AMSE, which allows a trade off between the bias and the variance.

Theorem 7. *The optimal smoothing parameter δ^* which ensures a pointwise trade off between the bias and the variance is given by $\delta^* = Cn^{-\frac{1}{2\alpha+3}}$, where the constant $C(F,\alpha,x)$ can be estimated separately. For this value of δ, we have* $\mathbf{AMSE} = O(n^{-4/2\alpha+3})$

Proof. Notice that $[\mathbf{Bias}]^2 = O(\delta^4)$ and $\mathbf{Var} = O(n^{-1}\delta^{1-2\alpha})$. Thus the order of MSE is $\max(O(\delta^4), O(n^{-1}\delta^{1-2\alpha}))$. So heuristically, the optimal value of δ will have to be such that the bias and variance have the same order of magnitude.

Let δ_n be a sequence of the form $kn^{-\beta}$. Therefore the asymptotic optimal choice of β (such that the two terms in the previous sum are of the same order) is given by

$$\begin{aligned} \delta^4 &= n^{-1}\delta^{1-2\alpha} \Leftrightarrow \\ n^{-4\beta} &= n^{-1}n^{-(1-2\alpha)\beta} \\ \beta &= \frac{1}{2\alpha+3} \end{aligned}$$

A more formal proof can be obtained by differentiating the AMSE with respect to δ_n.

$$\delta^* = \arg\min_{\delta_n} \mathbf{AMSE}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta)\right]$$

The first order condition is :

$$(1-2\alpha)n^{-1}\delta^{-2\alpha}\mathbf{F}^{(1)}(x)\frac{\mathbf{\Gamma}(2\alpha-1)}{[\mathbf{\Gamma}(\alpha)]^2} + 4\frac{1}{24^2}\alpha^2\delta^3\left[\mathbf{F}^{(\alpha+2)}(x)\right]^2 = 0$$

that gives

$$\delta^* = Cn^{-\dfrac{1}{2\alpha+3}}$$

where the constant C depends on α and the unknown values $\mathbf{F}'(x)$ and $\mathbf{F}^{(\alpha+2)}(x)$

$$C\left(\mathbf{F},\alpha,x\right) = \left[144\frac{(2\alpha-1)}{\alpha^2}\mathbf{F}^{(1)}(x)\frac{\mathbf{\Gamma}(2\alpha-1)}{\left[\mathbf{\Gamma}(\alpha)\right]^2}\frac{1}{\left[\mathbf{F}^{(\alpha+2)}(x)\right]^2}\right]^{\dfrac{1}{2\alpha+3}}$$

With this choice of δ we have:

$\mathbf{MSE}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(x,\delta^*)\right] =$

$=\left[Cn^{-\frac{1}{2\alpha+3}}\right]^{1-2\alpha} n^{-1}\mathbf{F}^{(1)}(x)\frac{\mathbf{\Gamma}(2\alpha-1)}{[\mathbf{\Gamma}(\alpha)]^2}+\frac{1}{24^2}\alpha^2C^4n^{\frac{-4}{2\alpha+3}}\left[\mathbf{F}^{(\alpha+2)}(x)\right]^2+o(\delta^4+ n^{-1}\delta^{1-2\alpha})$

$= C^{1-2\alpha}n^{\dfrac{-4}{2\alpha+3}}\mathbf{F}^{(1)}(x)\frac{\mathbf{\Gamma}(2\alpha-1)}{[\mathbf{\Gamma}(\alpha)]^2}+\frac{1}{24^2}\alpha^2C^4n^{\dfrac{-4}{2\alpha+3}}\left[\mathbf{F}^{(\alpha+2)}(x)\right]^2+o(n^{-4/2\alpha+3})$

So we have **AMSE** $= O(n^{-4/2\alpha+3})$ and the theorem follows. □

The choice of this measure implies that one wishes to estimate the derivative in a single point x. If the goal is to estimate the derivative over a range of values x than the minimization of the AMISE must be performed. This means that we must integrate the expression above and obtain the equivalent order with respect to the powers of δ_n.

We have the following result:

Theorem 8. *Let $\mathbf{F}^{(\alpha+2)}(x)$ be bounded continuous function that is square integrable. The following asymptotic expression holds:*

$$\mathbf{MISE}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(\cdot,\delta)\right] = n^{-1}\delta^{1-2\alpha}\frac{\mathbf{\Gamma}(2\alpha-1)}{\left[\mathbf{\Gamma}(\alpha)\right]^2}+ \\ +\frac{1}{24^2}\alpha^2\delta^4\int\left[\mathbf{F}^{(\alpha+2)}(x)\right]^2dx+o(\delta^4+n^{-1}\delta^{1-2\alpha})$$

The optimal choice of δ as a function of n is in this case:

$$\delta^* = Cn^{-\dfrac{1}{2\alpha+3}}$$

where the constant is in this case:

$$C\left(\mathbf{F},\alpha\right) = \left[144\frac{2\alpha-1}{\alpha^2}\frac{\mathbf{\Gamma}(2\alpha-1)}{\left[\mathbf{\Gamma}(\alpha)\right]^2}\frac{1}{\int\left[\mathbf{F}^{(\alpha+2)}(x)\right]^2dx}\right]^{\dfrac{1}{2\alpha+3}}$$

Again, with this choice we have

$$\mathbf{MISE}\left[\widetilde{\mathbf{F}}_n^{(\alpha)}(\cdot,\delta)\right] = C^{1-2\alpha}n^{-\dfrac{4}{2\alpha+3}}\frac{\mathbf{\Gamma}(2\alpha-1)}{\left[\mathbf{\Gamma}(\alpha)\right]^2}+ \\ +\frac{1}{24^2}\alpha^2C^4n^{-\dfrac{4}{2\alpha+3}}\int\left[\mathbf{F}^{(\alpha+2)}(x)\right]^2dx+o(n^{-4/2\alpha+3}) \\ = O(n^{-4/2\alpha+3})$$

There exists a large literature on choosing the bandwidth for the kernel estimator and all the proposed approaches might be adapted and used in this context.

7 Conclusion

We presented a theoretical estimator for the fractional derivative of the distribution function. There is no unique way to define the fractional derivative, and implicitly a nonparametric estimator, here we have chosen the Grunwald-Letnikov's definition, based on finite differences because it generalises an existing estimator for the case of integer order of differentiation. This definition is actually the one that is widely used in numerical computations for the fractional calculus, given its simplicity. Its bias, variance and convergence were studied. We found that the parameter δ can be seen as a "smoothing parameter", in the same way as the bandwidth h for the kernel estimator. We found that the convergence rate for our estimator with respect to δ, in the case of an integer α is the same as the convergence rate for kernel based density derivative estimation.

This estimator will allow us to perform nonparametric estimation as a useful tool for resolving some mathematical problems involving differential equations arising in finance, physics, dynamic systems, optimal controlled systems and others. Today, in many fields, theoretical models obtained for some integer value of k (order of differentiation) tend to be generalised to the fractional case, and this ask to review, in particular, the derivative of a functional parameter of interest in this context.

Given the multitude of the theoretical and numerical applications existing in the literature in the field of fractional calculus (fluid mechanics, material resistance, electrical behaviour of materials, optimal control, partial differential equations of fractional order etc.), we think that this work can be extended in different contexts that need a statistical study of the respective physical systems.

References

[1] R.J. Elliott and van der Hoek J. *Trends in Mathematics*, chapter Fractional Brownian motion and financial modeling., pages 140–151. Birkhauser Verlag Basel/Switzerland, 2001.

[2] G.M. Fichtenholz. *Functional Series*. Gordon and Breach Science Pub New York, 1970.

[3] T.T. Hartley and C.F. Lorenzo. Dynamics and control of initialized fractional-order systems. *Nonlinear Dynamics*, 29:201–233, 2002.

[4] R. Hilfer. *Applications of Fractional Calculus in Physics*. World Scientific, 2000.

[5] J. L. Knight. Moments of ols and 2sls via fractional calculus. *Econometric Theory*, 2:291–293, 1986.

[6] G. Laue. Remarks on the relation between fractional moments and fractional derivatives of characteristic functions. *Journal of Applied Probability*, 17(2):456–466, 1980.

[7] S.J. Lin. Stochastic analysis of fractional brownian motion,fractional noises andapplications. *SIAM Review*, 10:422–437, 1995.

[8] F Mainardi. *Fractals and Fractional Calculus in Continuum Mechanics*, chapter Fractional calculus: some basic problems in continuum and statistical mechanics, pages 291–348. A. Carpinteri and F.Mainardi, springer verlag, wien edition, 1997.

[9] C. Maltz. Estimation of the k-th derivative of a distribution function. *The Annals of Statistics*, 2(2):359–361, 1974.

[10] F.J. Molz, J. Fix, and Silong Lu. A physical interpretation for tha fractional derivative in levy diffusion. *Applied Mathematics Letters*, 15:907–911, 2002.

[11] X Moreau, C. Ramus-Serment, and A. Oustaloup. Fractional differentiation in passive vibration control. *Nonlinear Dynamics*, 29:343–362, 2002.

[12] K. B. Oldham and J. Spanier. *The Fractional Calculus: Integrations and Differentiations of Arbitrary Order*. New York: Academic Press, 1974.

[13] Emanuel Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[14] P.C.B. Phillips. The exact distribution of the sur estimator. *Econometrica*, 53(4):745–756, 1985.

[15] P.C.B. Phillips. The exact distribution of the wald statistic. *Journal of Econometrics*, 54:881–895, 1986.

[16] F. Riewe. Mechanics with fractional derivatives. *Physical Review*, 55:3581–3592, 1997.

[17] Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.

[18] S.G. Samko, A.A. Kilbas, and O.I. Marichev. *Fractional Integrals and Derivatives. Theory and Applications.* Gordon & Breach, 1993.

[19] E. Scalas, R. Gorenflo, and F. Mainardi. Fractional calculus and continuous-time finance. *Physica A Statistical Mechanics and its Applications*, 284:376–384, 2000.

[20] Vu Kim Tuan and R. Gorenflo. Extrapolation to the limit for numerical fractional differentiation. *Zeitschrift fur Angewandte Mathematik und Mechanik*, 75(8):646–648, 1995.